TS-EAS Subteam on Related Standards
Report 2018/2019
Submitted by Regine Heberlein, August 2019

Members
Kerstin Arnold
Regine Heberlein (chair)
Joost van Koutrik
Gerhard Müller
Nikitas Tampakis (Princeton University)
Donald Thornbury (Princeton University)

Repository
https://github.com/SAA-SDT/TS-EAS-subteam-notes/tree/master/related-standards-subteam

Charge and Activities
At the August 2018 Committee of the Whole meeting in Washington, DC, the Subteam on Related Standards was given the charge

> *to identify, investigate, and evaluate data models and ontologies that relate to the ongoing development of the Encoded Archival Standards (EAS) suite in terms of their alignment with and potential to inform the current trajectory of EAS.*

The subteam met eight times starting in September 2018. The group hosted one guest speaker, Elizabeth Russey Roke of the SAA Linked Data Interest Group, in February to talk about the work of the LD Interest Group.

Agenda items were selected collaboratively. The related standards and models varied considerably by type (conceptual model, ontology, communication standard, application profile, implementation etc.), application (discovery, description, delivery), and descriptive scope (encompassing all or a subset of the EAS suite):

- BIBFRAME and ARM (BIBFRAME extension)
- schema.org/architypes
- ISO 23081 (and ISO 30301 / 15489 / 26122 if applied in records systems)
- Europeana Data Model (EDM)
- MARC-21 authorities
- Functional Requirements for Bibliographic Records, object-oriented (FRBR-oo)
- Library Reference Model (LRM)
- wikidata/wikibase
- ArcLight implementation

Two members of the subteam, Kerstin and Regine, also explored the RiC ontology (RiC-O). Due to the ongoing development of RiC-O and the confidentiality requested by the RiC-O development team, RiC-O was not included in the discussions of the subteam as a whole.

Summary reports on the models discussed and supporting documentation may be found in the subteam's GitHub repository.

Based on their analysis, the subteam assigned high priority to ongoing discussions and investigations of FRBR-oo/LRM, wikidata, and EDM. The following are excerpts from the corresponding individual subteam reports:

EDM

The EDM has been developed as part of the various projects that brought to life Europeana, the cross-domain aggregator for cultural heritage. The portal includes digital/digitised objects from libraries, archives, museums and archaeological collections and hence EDM includes bits and pieces from the dominating standards in these domains (and others). There have been smaller specialist projects coming out of Europeana, e.g. on audio-visual content or the topic of fashion, which have reused EDM and extended it to fit their specific purposes. Europeana also collaborated with the Digital Public Library of America (DPLA) on the aspect of rights information.

In terms of EAD, Europeana's engagement with the Archives Portal Europe as the archives domain aggregator has led to various approaches of mapping EAD to EDM. The main challenge in this context has been the hierarchical approach of EAD including the concept of inheriting information from higher levels of description to the file/item level, which doesn't necessarily fit well with EDM's object-centric approach.

EDM's strengths and weaknesses are in its endeavour of bringing together data from - sometimes - very different domains and traditions. There are good aspects in the model with regard to e.g. breaking it down to DC elements for the main descriptive metadata and thereby agreeing on general terminology rather than accommodating "different terms for the same thing". On the other hand, this general view also comes with the need to compromise.

One area within EDM might be of interest for EAS, especially for EAD: rights declaration using Creative Common licences as well as some additional rights statements. These are now maintained by the Rights Consortium of Europeana and DPLA and others, i.e. they seem to be getting more attention. There might be a possibility there to (a) be more specific with regard to what's currently captured in <accessrestrict> and <userestrict> and (b) to perhaps be more interoperable.

The Archives Portal Europe has tried to integrate this information via a combination of EAD and METSRights, creating a specific profile of the latter.

wikidata/wikibase

Wikidata is a collaborative knowledge base implemented as linked data in the wikibase technology. Wikibase is a light-weight technology stack with the promise of getting linked data off the ground on a large scale. All it's missing is some descriptive rigor--that's where the GLAM community comes in.

Crucially, while wikidata can be viewed skeptically as a crowdsourced ontology of sorts, it has also come into its own as a "linking hub" that facilitates mappings and same-as and class relationships between ontologies, domain-specific vocabularies, and natural languages. Its greatest strength is its agility and ease of use. "Items" can be created instantly with minimal effort; "properties" go through an agile community approval process.

This flexibility is viewed as a liability in the traditionally very prescriptive context of GLAM resource description and must be mitigated by the additional observation of external conceptual models such as RiC or FRBR-oo. Shape expressions, formally introduced to wikidata on May 29 2019, allow validation of particular profiles. Open editing is viewed as another liability of the wikidata ecosystem. In a library context, it should be negotiated by the deployment of local institutional instances of wikibase that in turn exchange data with the open instance. This requires a set of policies for data sharing and data harvesting, respectively. It allows institutions to retain control of who edits their local records.

For archival description (among others), wikidata and wikibase are ready to go, with the startup investment described above, and would allow it the quantum leap into the realm of truly linked data.

FRBR-oo/LRM-oo

FRBR-oo is a conceptual model that combines FRBR, geared towards bibliographic items, and CIDOC, a museum standard geared towards artifactual objects. What's great about it is that it is a generalized model for resource description, rather than a format-specific one. It accommodates archival description almost out of the box, including mechanisms for describing entities such as data carriers and events in structured ways, where this is sorely lacking from EAS now. It is also linked-data ready.

The expectation is that following the release of the Library Reference Model (LRM), FRBR-oo will eventually be superseded by LRM-oo.

FRBR-oo includes some crucial entities that are currently lacking from the EAS standards, notably entities for carriers, events, and containers (the latter via a workaround class "Type" (E55)). These entities facilitate the description of content separately from physical properties and allow the incorporation of structured administrative data (such as curatorial events (exhibits, loans, accessions, conservation action) and pre-acquisition data (custodial history, archaeological data (findspot, excavation) into one system).

FRBR-oo is linked-data ready. Its relationships are published as URIs. It is also modular enough to be built out without having to start over (it would be fairly easy, for example, to add a container entity

without disturbing the overall consistency of the model). It allows multiple lateral relationships (e.g. several works coming together in one expression and/or manifestation). It is granular enough to accommodate a number of different types of aggregations (Complex Work; Container Work; Aggregation Work; Serial Work). And it greatly improves FRBR by adding production layers to the WEMI structure in the form of Production Expression and Manifestation Product Type.

Challenges for EAS

The inclusive charge allowed the subteam to explore and discuss a variety of conceptual data models, ontologies, and implementations. The group prioritized two discussion points: the transition of archival description to a linked data environment, and the need for a robust conceptual data model to facilitate that transition.

The current suite of XML-based EAS is rooted in a long-standing narrative tradition that doesn't easily translate to a linked data environment. The subteam identified and discussed the following issues relating to the transition of archival description to linked data:

EAS does not explicitly spell out its place in what DACS describes as the

> *cascading workflow of resource description that emanates from a system-agnostic structure standard, whether implied or explicitly stated, and proceeds through a set of content guidelines to its final implementation by ways of a communication standard.* (Revised preface)

This results in EAS taking an independent approach in the ecosystem of GLAM resource description that could be more harshly seen as isolationist. The following two issues in particular underscore this point:

- In the absence of an explicit conceptual data model, EAS serves double duty as an implicit data model. EAS is built on dated monolithic concepts such as "collection" and "descriptive record," among others, that inform its implicit data model in ways ill suited to a transition to linked data. For example, the primacy of the collection-based record creates relatively hermetic information packages in which individual resources are represented as subordinate to, rather than parts of, a collection, and cannot easily be represented as belonging to more than one collection.
- TS-EAS does not explicitly define EAS[1] as an implementation of DACS or any other content standard. This agnosticism is a deliberate accommodation to allow EAS to serve an international community. It has unintended consequences, however, that present barriers to a transition to linked data. Most notable among those is a data permissiveness that makes EAS both hard to maintain and hard to transform into a uri-driven environment. Consider for example ongoing discussions about value validation for identifier codes.

[1] Currently comprised of Encoded Archival Description (EAD), Encoded Archival Context--Corporate Bodies, Persons, Families (EAC-CPF), and an emerging Encoded Archival Context--Functions (EAC-F). A subteam of TS-EAS is currently working towards aligning the EAD and EAC-CPF schemas.

A Revised Vision and Mission for TS-EAS

In recognizing the above challenges, the discussions of the Subteam on Related Standards therefore included bigger-picture questions relating to the mission and vision of TS-EAS and the subteam's role within those parameters. These questions can be summarized as follows:

1. The purpose of TS-EAS as currently stated includes two charges: "ongoing maintenance" of the current suite, including development of future companion standards; and "thoughtful evolution of archival encoding standards on behalf of and in consultation with the global archival community." The current work of TS-EAS is heavily slanted towards the former. It might be helpful for TS-EAS to formulate a more detailed mission statement relating to the latter point to articulate its role in driving the evolution of archival description.
2. The purpose of TS-EAS as currently stated suggests that TS-EAS works "in close collaboration with the International Council on Archives Experts Group on Archival Description." EGAD is the group currently developing RiC, yet discussions in TS-EAS as a whole during 2017-19, and assertions made by the co-chairs during that time, suggest that TS-EAS is reserving the right to pass on RiC at its discretion, and no concerted efforts have been made to align the two. It might be helpful therefore for TS-EAS to articulate in more detail the role of EAS vis-à-vis work being done at ICA, in particular whether EAS is to be an implementation of RiC-CM. If it is determined that it is not, TS-EAS may further consider its position towards adopting a different CM or taking responsibility for formulating one; in either case documentation of that position will help future members of TS-EAS, as well as the professional community, in their work.

Recommendations

Having surveyed a broad spectrum of data standards, models, and implementations, the subteam identified a small number of candidates, which warrant a more detailed investigation, and recommends the following action items:

1. Immediate action items
    - Form a project team to develop a set of actionable **use cases**.
    - Conduct a **comparative evaluation of RiC-CM and, pending the release of LRM-oo, FRBR-oo** (which has a broader scope and the structure of which informs RiC) with the goal of identifying a robust conceptual data model developed for a linked data environment.
    - Conduct comprehensive **testing of wikidata and wikibase** (in conjunction with a CM and/or shape expressions) as a pathway to archival description as linked data.
    - Conduct a **cost/benefit analysis of "retrofitting" EAS** with uri's v. RiC-CM and RiC-O v. remodeling EAS from scratch for linked data.

2. Prospective action items

- **Maintain a running overview of emerging related standards** and keep it current to serve as a reference tool for members of EAS.
- Work towards strengthening the **alignment of EAS with content standards** such as, to name examples without further recommendation, ISAD(G), DACS, AMREMM/DCRM (soon to be merged into the RDA policy statements) or similar, with particular attention given to the chosen standards' evolution towards linked data, e.g. DACS Principle 4.[2]
- **Forge collaborative alliances** with stakeholder communities of related standards to work towards coevolution of related standards with EAS. For example, involve the SAA Records Management Section in any further work relating to ISO records management (metadata) standards and encoded archival description standards.

[2] "Records, agents, activities, and the relationships between them are the four fundamental concepts that constitute archival description."
https://github.com/saa-ts-dacs/dacs/blob/master/statement_of_principles.md#4-records-agents-activities-and-the-relationships-between-them-are-the-four-fundamental-concepts-that-constitute-archival-description